And it would appear that many of them exist just because the comprehension device is just the sort of heuristic information processor we have suggested. Without such constraints, misunderstanding would be endemic. Bever and Langendoen (1971) have presented an extended discussion of this point in the context of constraints on relative clause formation as they have emerged through the history of English.

We have established informally that decisions concerning structure and lexical readings are not delayed indefinitely long awaiting disambiguating context. Specifying when and how such decisions are made is considerably more difficult. In some sense, the question is whether there is a point, defined either temporally or structurally, at which such decisions are characteristically made. There might be, of course, several such points, differing in their location and in their determination, for different kinds of decisions. For example, it is possible that decisions about readings for individual lexical items are made earlier and on the basis of less intensive information than those about major structural segments, e.g., clauses.

Consider again the garden-path sentences (3) - (6) discussed earlier. One feature they share is that the material likely to be misinterpreted is separated by a clause boundary from the material specifying the correct interpretation. This feature, which is common to most such sentences, suggests that decisions about structure and readings for lexical items are characteristically not deferred past the end of the clause in which they occur, for if they were, the disambiguating material would have been encountered before an interpretation was assigned to the garden-path portion of the sentence. More exactly, it appears that such decisions are made at the point at which the hearer has sufficient information available to determine that the clause has ended.[6] Generally, this is a point somewhat past the objective location of the boundary, though in some cases it may actually precede the objective boundary (Hakes, 1971b).

There is, in addition, a considerable body of research which suggests that the clause is a primary unit of comprehension processing. For example, research on the perception of "clicks" presented during sentences

---

5 (con't) verb's possible readings is available immediately on processing the verb and prevents, in the case of look but not in the case of call, assigning a reading before the particle is encountered. That is, the difference in acceptability seems to arise from the range of readings possible for the verb even though they are not realized. Unfortunately, consideration of other examples would show that the correct account must be even more complicated than we have suggested.

[6] It is at this point quite unclear how this constraint should be stated. We have stated it in terms of the end of a clause, but it seems entirely possible that the beginning of a clause might have the same effect even if the preceding clause were not ended. Essentially, this is a question about whether the boundary between believe and John in [Mary believed [John was a fool]] is the occasion for assigning structure to [Mary believed it].

suggests that such clicks tend to be displaced perceptually toward clause boundaries (e.g., Bever, Lackner & Kirk, 1969; Fodor, Bever & Garrett, in press). Bever, Garrett and Hurtig (in preparation) have shown, using a sentence completion task, that structural ambiguity has an effect on completion latency only if the fragment containing the ambiguity is less than a full clause. This suggests that once the clause has been completed a structure is assigned (i.e., that after that point there is no more processing to be done for an ambiguous fragment than for an unambiguous one). However, the same study showed that completing a lexically ambiguous fragment took no longer than an unambiguous one, regardless of whether the fragment was a full clause or not, suggesting that decisions about lexical readings may be made earlier than those about grammatical structure.

Caplan (1972) has found similar clause boundary effects using a probe task in which the S hears a sentence followed by a single word and must decide as quickly as possible whether the word occurred in the sentence. Essentially, Caplan found that if the target (i.e., the word being probed) was in the sentence's last clause, response latency was shorter than the target was in an earlier clause. Since the distance between the target and the probe was equated across these conditions, the results suggest that lexical material in the last clause is more readily available than earlier material.

Thus, there seems to be considerable evidence that decisions about underlying grammatical relations and readings of lexical items are generally not deferred past the end of the clause. It may be, however, that how long a decision is deferred is at least partially under the hearer's control. While there is no "hard" experimental evidence to this effect, common gossip has it that performing experiments on garden-path sentences is extremely difficult for just this reason. That is, once a S discovers that he is being garden-pathed, he is no longer fooled by such sentences. What apparently happens is that he defers making decisions longer than he normally would in the anticipation that disambiguating information will be forthcoming. The extent to which this can be done is probably influenced by the "cost" of deferring. That is, if S is being presented isolated sentences, e.g., in an experiment, he may be better able to do this than if he is listening to continuous discourse, a situation in which the memory load incurred by deferring decisions would probably become intolerable.

The fact that structural and lexical decisions are generally not deferred past the end of the clause does not, of course, imply that all such decisions are made at the clause boundary. The Bever, Garrett and Hurtig experiment described above suggests that at least some decisions about lexical readings are made prior to the clause boundary. In addition, it is possible to construct some garden-path sentences which exhibit this, e.g., sentence (9). Further, since sentence-internal clause boundaries are generally not marked in the acoustic signal, the comprehension

(9) The fat grocer weighed two hundred grapes.

device must do at least enough processing within a clause to determine when it has ended (see Hakes, 1971b, for a discussion of the minimal

syntactic and lexical processing necessary to accomplish this). The question which remains is, then, how much more than this minimal processing occurs during a clause. We will return to this question below.

Up to this point we have been concerned primarily with some characteristics of making decisions about a sentence's underlying grammatical relations and readings for its lexical items and have said little about the possible bases for such decisions. We now turn briefly to that question.

To take a limited example, consider how the comprehension device might determine the internal structure of a NP with prenominal adjectives, e.g., (10). Much discussion of prenominal adjectives and their ordering

(10) The small green square block...

has recently appeared in the literature (e.g., Vendler, 1968; Martin, 1969, 1970; Bever, 1970). It is generally agreed that a lexical item that can appear as a prenominal adjective can also appear as a noun, though not necessarily the converse. Much of the discussion of adjective ordering has focused on characterizing the normal ordering of prenominal adjectives, there being little agreement on the basis for the ordering. There is agreement that the ordering is from least to most "nounlike", though it remains unclear whether the basis for ordering is semantic, syntactic or some combination of the two.

Suppose that the comprehension device, in reading through a sequence such as (10), immediately looks up each item in its dictionary to determine (at least) its possible form class assignments. It seems unlikely that information about "nounlikeness" would be of much use here, for except in the case in which the last item was unambiguously a noun, it would always be possible for there to be another item following that which was still more nounlike. A more plausible strategy would be to look for the first item in the sequence that could <u>not</u> be a noun. Upon finding such an item, it is virtually certain that the item preceding that item is the head noun of the NP and that the items preceding it are adjectives. That is, there is a constraint on English to the effect that the head noun of a NP cannot be followed immediately by another noun. Put differently, it is not the fact that the last item within a NP is the most nounlike that is important; rather it is the fact that it is last. The fact that NPs are occasionally missegmented provides further evidence that the correct strategy is based on "overshooting" the end of the NP, for the cases that are prone to missegmentation are exceptions to the constraint that a noun cannot immediately follow the end of an NP. Thus, sentence (11) tends to be misheard as being about green golf balls.[7]

(11) The green golf balls roll on best is a level, well-trimmed one.

---

[7]The fact that (11) is a rather atrocious sentence merely indicates that the constraint in question is a strong one; exceptions are hard to find. Note that the kind of strategy described here must be at least partially

Bever (1970) has discussed a number of the possible strategies which the comprehension device might use for retrieving underlying grammatical relations from the superficial characteristics of a sentence. The emphasis of his work has been on the formal character of the strategies themselves rather than on the cues which trigger them or on when or how the comprehension device might use them. Thus, for example, he argues that the most generally applicable strategy, and the one first acquired by children, is essentially:

> Segment together any sequence X...Y, in which the members could be related by primary internal structural relations, "actor action object... modifier."

Put differently, if the sequence contains N...V...N, then other things being equal, the sequence represents the underlying grammatical relations, subject-verb-object.[8] "Other things being equal" may be taken as indicating that the sequence contains no cues which mark the sequence as representing an exception to this strategy.

If, as Bever suggests, the NVN strategy is the most generally applicable one, as well as being the first acquired, it seems reasonable to treat all other strategies as being exceptions to the general rule, marked by their own patterns of superficial cues. Thus, for example, if the verb in an NVN sequence is surrounded by was...-ed...by, the sequence is marked as reversing the usual underlying syntactic relationships, i.e., as being passive rather than active.

Since our major concern here is with the cues themselves and with the ways in which the comprehension device uses them, we will not consider Bever's suggestions further. The question to which we wish to turn is that of determining what cues the device actually uses and how it uses them. Clearly, the problem that arises with garden-path sentences, i.e., systematic misunderstanding, provides considerable evidence on this. Such evidence is, however, insufficient, for the interaction of the device's strategies with constraints on possible sequences

---

7 (con't) language-specific; this particular strategy could not work for a language like French which allows adjectives to occur both pre- and post-nominally.

[8] The form in which Bever states the strategy is, of course, misleading in implying that Actor-Action-Object is the general semantic relation underlying such sequences (cf., Fillmore, 1968; Katz, 1972), for such a sequence can reflect any of a number of underlying "case" relationships, depending upon the character of the lexical items the sequence includes. Two options exist for extending Bever's strategy. First, to make the strategy sensitive to the character of the lexical items and modulate the relations assigned accordingly. Secondly, to state the ouput of the strategy in syntactic rather than semantic terms, permitting a component of the comprehension device that operates on the output of the strategy component to provide the additional semantic characterizations. At present, there is no clear basis for a choice.

in the language results in very few such errors being made. That is, there are simply too few sentences that produce the garden-path effect. It thus becomes necessary to seek other, more subtle cases and other kinds of evidence.

In general, it seems likely that the difficulty encountered in comprehending a sentence provides an index of the way in which the comprehension device is functioning. Further, it seems likely that ease or difficulty of comprehending will be determined by the particular cues a sentence exhibits and by the relationship between these cues and the sentence's structure and meaning. Hakes (1969), summarizing suggestions made earlier by Fodor and Garrett (1967) and Fodor, Garrett and Bever (1968), has suggested that there are three general sources of difficulty in the cue-structure relationship. First, a sentence should be more difficult to comprehend if a cue which might have been present is not, i.e., an optional cue is deleted. Secondly, difficulty may arise from the fact that the correlations between cues and structures are less than perfect. That is, difficulty will arise if the cues are, in the sentence under consideration, associated with a structure other than the one with which they are customarily associated. Thirdly, difficulty may arise if there are many structures associated with a particular set of cues, i.e., the cues themselves are ambiguous.

We turn now to an examination of research relevant to these three cases.

Research on Optional Cues.

Most of the research on this question has centered on relative clauses. Generally, such clauses are introduced by relative pronouns, e.g., (12). The pronouns may, however, sometimes be deleted without altering the underlying structure or meaning of the sentence, e.g., (13). If such pronouns are clear cues to the sentence's underlying structure, then deleting them should force the hearer to rely on "secondary" cues, and consequently the sentence should be more difficult to comprehend.

(12) The car which the man that the dog bit drove crashed.

(13) The car the man the dog bit drove crashed.

Some intuitive support for this claim comes from the fact that English does not permit relative clauses to be reduced (i.e., the pronouns to be deleted) if the resulting sequence is highly likely to be misleading. The pronoun may be deleted if it is the object of the relative clause that's pronominalized but not if it is the clause's subject. Thus, for example, the relative clause in (14) cannot be reduced. The reason is obvious. Deleting the relative pronoun leaves a sequence, The girl was running down the street, which is most likely

(14) The girl (who) was running down the street fell and broke her leg.

to be interpreted as a single, main clause. That the constraint works

less than perfectly is indicated by sentences like (6), cited earlier.

(6) The horse raced past the barn fell.

Here it is the object of the relative clause that has been relativized. But, in addition, the clause has been passivized. The combination of these two operations leaves the verb of the relative clause adjacent to the subject NP of the main clause, leading to the same kind of confusion that would result if subject relatives like (14) could be reduced (see Bever & Langendoen, 1971, for further discussion of these cases).

The cue deletion hypothesis, suggested originally by Fodor and Garrett (1967), has been tested a number of times using the deletion of relative pronouns. Fodor and Garrett performed a series of experiments in which Ss were presented with doubly self-embedded sentences like (12), which differed only in whether the relative pronouns were present or deleted, the Ss being required to paraphrase each sentence. The data were analyzed for both paraphrase accuracy and latency, and for both measures performance was worse when the pronouns were deleted. The basic experiment was repeated five times, each repetition varying in ways intended to rule out alternative explanations; all repetitions yielded essentially the same results.

Hakes and Cairns (1970) extended Fodor and Garrett's research, using the sentences from their experiments. The Ss were required both to paraphrase and to monitor each sentence during presentation for the occurrence of a specified word-initial phoneme. Response latency in this phoneme monitoring task had previously been shown to be sensitive to at least some aspects of on-line comprehension processing (e.g., Foss, 1970; Foss & Lynch, 1969). Hakes and Cairns found that paraphrasing was worse and monitoring was slower when the relative pronouns were deleted. Similar results were obtained by Hakes and Foss (1970) for both paraphrasing and phoneme monitoring with a different set of doubly self-embedded sentences.

The results of these experiments indicate fairly unambiguously that deletion of an optional cue does increase comprehension difficulty. Equally important, the fact that deleting relative pronouns had this effect strongly implies that the pronouns themselves are important cues when they are available. The exact manner in which they are used, of course, remains unclear. We may speculate, however, that they serve to indicate clearly that the first NP in a sequence is in a different clause than the second.[9] Alternatively, we may say that the pronoun, signals the existence of a clause boundary between the NPs. The existence of such a boundary is less clearly marked in the absence of the pronoun in that a sequence NP NP ... may also be the beginning members of a conjoint NP that continues ... and NP. Blumenthal (1966) reports

---

[9]It would be more accurate to say the first is both in the same clause as the second and a different one, serving different grammatical functions in the two.

that Ss attempt to paraphrase the initial NP sequences of triply and quadruply self-embedded sentences as conjoint NPs, though we have found no evidence of such a tendency in our own research.

Additional evidence of the effects of optional cue deletion comes from two experiments by Hakes (1972) on deletion of the optional complementizer that. Many English verbs can take as direct objects either a simple NP or a full sentence--a NP predicate complement construction (see Rosenbaum, 1967). There is a large number of such complement constructions but the one used in the experiments was the that-complement, as in sentence (15). The complementizer that may often, though not always, be deleted without altering the underlying structure or meaning of

(15) The blind student felt (that) the material in the art course would be too difficult for him to understand.

the sentence. The difference between the reduced and unreduced cases is that, in the unreduced case, there is a cue immediately following the verb that indicates nearly unambiguously that the verb is intended in its complement sense.[10] In the reduced case, such information is not available until after the NP, in this case, until the verb would is processed. Thus, in the reduced case there is a longer period during which the NP following the verb is interpretable either as the verb's direct object or as the surface subject of an embedded complement sentence.

In both experiments, Hakes used sentences like (15) in both reduced and unreduced forms. The Ss were again required to perform both paraphrasing and phoneme monitoring tasks, the monitored phoneme occurring within the NP serving as subject of the embedded complement, e.g., the /m/ in material in (15). The phoneme monitoring data of both experiments strongly supported the cue deletion hypothesis, i.e., monitoring latency was substantially longer when the complementizer was deleted. However, only one of the two experiments yielded a significant effect in the paraphrasing task. Hakes argued that the sentences in the other experiment were too easily comprehended, with or without the complementizer, to show a difference in difficulty in that task.

Thus, it would appear that the deleterious effect of deleting optional cues is general and that both relative pronouns and complementizers are important cues when they are available. Unfortunately, it is very difficult to extend this line of research to yield further information about the nature of the cues used during comprehension. The reason is just that it is very difficult to find cases in which deleting a potential cue has no effect on the sentence's structure or

---

[10]We say nearly unambiguously since, taken by itself, that is ambiguous between readings as a complementizer and as a demonstrative adjective or article, as in that boy standing over there... What this implies is that at least some information about the following NP is necessary to determine that that is in fact a complementizer.

meaning.

Linguists frequently argue that transformations, including deletion, do not effect a sentence's structure or meaning (see, e.g., Katz & Postal, 1964; Chomsky, 1965; Katz, 1972). But it should be made clear that this claim is intended in a sense different from that intended above. For it appears that the linguists intend the meaning-preserving claim as pertaining to individual sentences, i.e., deleting structure from a sentence does not alter the structure of that sentence. To take a well-worn example, consider the relationship between full and short passives, e.g., (16). The claim is that in both cases there is an underlying "agentive" NP.

(16) a. The boy was hit by the girl.
b. The boy was hit.

But what the linguists do not claim is that the particular lexical character of that NP can recovered from a short passive. Thus, (16a) and (16b) are only partially synomymous in that the truth conditions for the former include those for the latter but not conversely. Essentially, a lexically interpreted deep structure may include "dummy" lexical items, i.e., lexical items that are less than completely specified. It is just in case the structure contains such a dummy item in agentive position that a passive may undergo agent deletion. Thus, the structure underlying (16b) is preserved across transformations just because the agent never was specified (see Watt, 1970, for an extended discussion of this case and its implications).

But on this account (16a) differ just in that there is less information to be recovered about (16b) than about (16a). Thus (16b) might be easier to comprehend than (16a) for this reason. That is, the cues which are "deleted" are not really cues to underlying content, i.e., a fully specified agent. This seems to be a general feature of cases involving cue deletion; it is very difficult to find cases in which pairs of sentences can be constructed, with and without deletion, for which one can be confident that the same underlying structure and content must be recovered for both. Consequently, the search for cues must turn elsewhere.

Research on lexical complexity.

The hypothesis that the number of structures associated with a lexical item affects comprehension difficulty was first suggested by Fodor, Garrett and Bever (1968). The hypothesis originated in a model of comprehension processing holding that, on encountering a lexical item, the hearer retrieves from his mental dictionary information about the structures and meanings with which that item could be associated. This information is used to construct a set of hypotheses about the structure underlying the sentence being processed, hypotheses which are then tested against further information as it becomes available.

Arguing that structures associated with verbs are of particular importance, Fodor, Garrett and Bever focused their investigation on verbs

that differed in the number of different structures associated with them, choosing transitive and complement verbs. A transitive reading is one in which the verb takes simple NPs as both subject and direct object, where by "simple" we mean that the NP does not contain an embedded S. Thus, a "pure" transitive verb, such as kidnap, is one having no other readings and, as such, is a simple verb. Complement verbs, on the other hand, may take either simple NPs or sentential complement constructions as subjects and objects. Since there are a number of different complement constructions and since most verbs which can take such constructions can also take simple NPs, complement verbs are relatively complex. Note, of course, that the complexity does not exist within any given sentence, for on any given occasion a verb must appear in only one of its possible structures. The complexity in question is potential complexity.

Fodor et al reported two experiments testing the verb complexity hypothesis. The first used doubly self-embedded sentences which differed only in whether they contained a pure transitive or a complement verb in one of their embedded clause, e.g., (17) and (18). In all cases, the

(17) The letter the secretary the manager employed mailed was late (transitive)
(18) The letter the secretary the manager employed expected was late. (complement)

structure required a transitive reading. Subjects heard the transitive or complement versions of 12 such sentences, and were required to paraphrase each sentence on five successive presentations. The results indicated that paraphrasing was more accurate for the transitive verb versions.

In their second experiment, simpler sentences, again differing in whether they contained a transitive or a complement verb, were used in an anagram task. Subjects were presented the words of a sentence on individual slips of paper and were required to form the words into a sentence. No difference in time taken to complete the sentence was found, but the predicted difference was obtained both for failures to complete the sentence within 60 sec. and for incorrect solutions, i.e., non-sentences.

Hakes (1971a) reported two experiments designed to examine the verb complexity hypothesis further. In the first, paraphrasing and phoneme monitoring data were collected for pairs of sentences like (19) and (20) which differed only in whether the verb of the sentence's main clause was a pure transitive or a complement verb. The pairs of verbs were matched for frequency of occurrence and for initial phoneme and length.

(19) When he had heard all sides of the dispute, the manager blamed the foreman who had supervised the job. (transitive)
(20) When he had heard all sides of the dispute, the manager believed the foreman who had supervised the job. (complement)

The monitoring target was either the verb itself or the head noun of the verb's object NP. The paraphrasing data replicated the effect reported by Fodor et al, i.e., paraphrasing was less accurate when the sentences

contained complement verbs that when they contained transitive verbs. However, the monitoring data yielded no difference as a function of verb type for either target location.

In the second experiment, similar pairs of sentences were used, differing from those of the first experiment in that location of the clause containing the verb occurred either early or late in the sentence. The target for the phoneme monitoring task was always the initial phoneme of the head noun of the object NP. Again, no effect of verb complexity was found in the monitoring data, regardless of whether the verb occurred early or late. At this time the paraphrasing data also failed to yield a verb complexity effect.

There are, of course, several possible accounts for this pattern of results. Two of three experiments yielded significant effects with the paraphrasing task; neither of two experiments yielded an effect for the phoneme monitoring task. First, it should be noted that there are reasons for believing that the paraphrasing task is a relatively insensitive measure of whatever it is that it is measuring (Hakes, 1972). Thus the failure of one of the three experiments to yield a significant verb complexity effect for the paraphrasing task may simply reflect the task's insensitivity rather than the lack of such an effect.

A more serious problem concerns whether the paraphrasing task reflects difficulty in comprehending a sentence or difficulty in some other process intervening between the input of a sentence and the output of a paraphrase. Suppose, for example, that the subject comprehended the sentences of these experiments perfectly, regardless of whether they contained transitive or complement verbs, i.e., that the complexity of the complement verbs did not cause any greater processing load or any more comprehension errors. The subjects' further task is to take a representation of the sentence as comprehended, modify either its lexical content or its structure without altering its meaning, and produce the modified sentence. Suppose now that in modifying the sentence the subject used the main verb as a guide to changing structure, i.e., kept the same verb and selected a structure indexed by that verb. Since there is only one structure compatible with a pure transitive verb, this should not be a source of error for sentences containing such verbs. For a complement verb, however, there is some possibility of the subject's choosing a structure for the paraphrase different from that represented in the input sentence, failing to notice that the result was a change in meaning. The point is simply, that there may be greater possibility for error in generating a paraphrase for a complement verb sentence than for a transitive verb sentence and, hence, that the obtained difference in paraphrasing accuracy might reflect this rather than a difference in comprehension difficulty.

Such an account fails, however, on other data. Hakes (1972) has pointed out that the paraphrasing data of the experiments on relative pronoun and complementizer deletion are not subject to this criticism. For those sentences it must be assumed that the structure and content retrieved is the same (i.e., if comprehended perfectly) regardless of whether the pronoun or complementizer is present or absent. That is, the representation on which a subject may be assumed to base his para-

phrase is the same. Since these experiments yielded differences in paraphrasing which cannot be attributed to differences in post-comprehension processing, it follows that the differences must be attributable to differences during comprehension. Arguing from like effects to like causes, it follows that the differences obtained as a function of verb complexity are also attributable to differences in comprehension difficulty.

Thus, it seems likely that the paraphrasing effects obtained in the verb complexity experiments are truly effects of comprehension processing. But if so, the question remains of why the effect was not reflected in the phoneme monitoring data. There are numerous questions about whether the phoneme monitoring task reflects processing load during comprehension and, if so, what aspects of comprehension processing are reflected, and we will return later to raise these questions briefly. But for the moment let us assume that the task does reflect comprehension difficulty. Why,then, was no verb complexity effect obtained in either experiment?

The model on which the experiments were designed was that proposed by Fodor et al and holds that the hearer projects hypotheses about a sentence's possible predicate structures on the basis of stored lexical information about the verb. Certainly, if this is the case, then the reasonable place to look for an effect of verb complexity would be after the verb but before sufficient information was available to determine uniquely the correct structure. For the sentences in both experiments, since the structure was in fact always that of a verb followed by a noun phrase, there was sufficient information available at or shortly after the end of that NP to determine that the correct reading for the verb was just where the two experiments failed to find such an effect. Consequently if we are to retain the assumption that the monitoring task is sensitive to the kinds of effects sought, we must conclude that the Fodor et al account is wrong.

Hakes (1971b) has presented this argument in detail and, in addition, suggested an alternative model. He noted that it would not be particularly economical for the hearer to make full use of information about possible predicate structures immediately since the question of which structure is actually realized is almost always resolved shortly after the occurrence of the verb, i.e., by the time the following clause boundary has been identified. For this and other reasons Hakes suggested a model which holds that during a clause the hearer does little more than try to determine the major form class assignments of the items encountered. This information (plus information about the sequence in which the items occur) is sufficient to permit the hearer to discover a clause boundary when he encounters one. Since the question of the clause's internal structure is generally resolved at that point, Hakes suggested that it is not until then that the hearer attempts to assign structure to the clause. Hence, it should not be until that point that a difference in verb complexity should manifest itself.[11]

---

[11]This "passive" processing model incorporates some rather complex assumptions about the way in which dictionary information about lexical items is stored and retrieved. For example, it requires the assumption that the hearer can determine whether some item has a reading as a verb

Based on these assumptions, Hakes performed two further experiments looking for on-line processing effects of verb complexity. For both experiments pairs of sentences were constructed that differed only in whether they contained a transitive or a complement verb. In all cases that verb was followed by a simple NP which, in turn, was followed by a clear clause boundary, as in sentence (21) and (22). The monitoring target

(21) After opening them, the purchasing agent reviewed the bids, but none was low enough to be accepted. (transitive)

(22) After opening them, the purchasing agent revealed the bids, but none was low enough to be accepted. (complement)

was located shortly after the clause boundary, for these sentences, the /n/ in none.

The paraphrasing data for the first of the two experiments did not yield a significant verb complexity effect. However, for the second experiment, in which the sentences were more complex, the complement verb sentences were paraphrased significantly less accurately than those containing transitive verbs. In neither experiment was there even a hint of a verb complexity effect for the phoneme monitoring task.

The phoneme monitoring data for the four verb complexity experiments, taken together, indicate that no processing load effect of verb complexity manifests itself either before or after the end of the clause. In light of these data Hakes (1971b) extended the "passive" processing model. He noted that the data could be interpreted as indicating not only that retrieving and storing information about a complex verb "costs" no more than for a simple verb but, in addition, that deciding that the structure is transitive costs no more when the verb has other readings than when it does not. That is, given that the structure to be assigned is the same regardless of whether the verb has other readings, it may be that the assignment of that structure can be made without considering the other readings, i.e., without their interfering.

If the structural decision is made just after the clause boundary, then at the point of decision there is sufficient information available to determine that the relevant structure is $[\ldots V\ NP]_S$, i.e., that the verb is followed by a simple NP, followed by the end of the clause. If the search for a reading for the verb were conducted at that point, it might be possible to search only for a reading that fit that context. Thus, if the search code was context-sensitive, containing a specification that the verb occurred in the context of a simple NP, then essentially the only question that need be asked is, "Does this verb have a transitive reading?" For lexical information stored in a content-addressable

---

11 (con't) without at the same time retrieving other information about that item. Some aspects of the model will be discussed later. But suffice it to note here that the assumptions about lexical storage and retrieval could be largely correct even if the hypothesis that structure assignment is made at and only at a clause boundary is false.

memory, this question should be no more difficult to answer for a complex verb than for a simple verb. Hence, if the context between the verb and the following clause boundary uniquely specifies the appropriate reading, verb complexity should have no effect on processing load, regardless of where processing load is examined.

This account is, of course, based almost entirely on negative evidence, on the failure to find phoneme monitoring effects where anticipated. Further, it presents considerable difficulty in accounting for phoneme monitoring data collected in other situations as well as in accounting for the significant paraphrasing effects obtained in verb complexity experiments. Nonetheless, it does make an interesting prediction, one worth examining even if the general account is false.

The passive processing model suggested here assumes that structural decisions are made immediately following clause boundaries. Further, in sketching the model we assumed that by the time a clause boundary is crossed there is sufficient information to allow an unequivocal structural decision. That this is not always the case is clear from a consideration of ambiguous sentences, particularly the garden-path sentences discussed earlier. Suppose now that we were able to find a case in which the NP following the transitive or complement verb was followed by a clear clause boundary but where, in addition, there was not at the clause boundary sufficient information to determine whether the NP preceding the clause boundary was the direct object of the verb or the subject of a predicate complement. A case of this sort is presented in sentence (23a). Here the sentence could end as in (23b), where the beginning of the relative clause is, in fact, the end of the preceding clause, and the verb suggested has a transitive reading with several reforms as direct object. The sentence could, however, end as in (23c), where the relative clause merely interrupts the complement sentence. But the hearer could not know which kind of completion for (23a) was correct until after the end of the relative clause, for it is not until that point that he could discover whether or not there is a predicate phrase to be associated with the NP, several reforms.

(23) (a) The committee's report suggested several reforms, which...
(b) ...were badly needed, though no one expected them to pass.
(c) ...were badly needed, should be passed immediately.

There are two parallel cases for which this problem does not arise. First, if the sentence contained a pure transitive verb, e.g., contained, rather than the complement verb suggested, then (23a) would have to be completed with the structure of (23b) and, further, there would be sufficient information available at least by the beginning of the relative clause to determine this. Second, the complement verb in (23a) could be followed immediately by the complementizer that, in which case the only possible structure would be that of (23c) and, again, information would be available by the beginning of the relative clause to determine this. Put differently, the local ambiguity created by the verb is, for these latter two cases, resolved almost immediately, while in the first case the ambiguity remains unresolved for considerably longer.

There seem to be two kinds of strategies the hearer could adopt in such cases. First, he could assume that the beginning of the relative clause does not signal the end of the preceding clause and defer assigning structure at least until the end of the relative clause. If he were to do this, it would be necessary to store a large amount of relatively unanalyzed material, a fact that should be reflected in monitoring latency to a target located within the relative clause. But this strategy seems unlikely, in part because it would require storing a large amount of material and in part because such clause boundaries usually do signal that the preceding clause has ended.

The second, more likely strategy is that the hearer could assume that the beginning of the relative clause did not signal the end of the preceding clause and attempt to assign structure at that point. In this case, there would not be an abnormally long latency for a monitoring target within the relative clause. But because two structures are possible, the hearer might well assign the wrong structure. Thus, if he assumed the structure was transitive, he would be wrong just in case the sentence ended like (23c). And if he assigned the complement structure, he would be wrong just in case it ended as in (23b). In either case, he stands in danger of being garden-pathed.

To examine this situation, Hakes performed another verb complexity experiment, collecting paraphrasing and phoneme monitoring on each of five variants of a single sentence type, exemplified in (24). (24a) contains a pure transitive verb. In (24b), the verb occurs in its transitive reading, and the sentence is completed in the same way as (24a). (24c) and (24d) both involve the complement reading of the verb and differ only in that (24c) contains the complementizer <u>that</u> while (24d) does not.

(24) (a) After investigating the accident, the manager <u>blamed</u> the foreman, who had supervised the job, but decided not to take any action.

(b) After investigating the accident, the manager <u>believed</u> the foreman, who had supervised the job, but decided not to take any action.

(c) After investigating the accident, the manager <u>believed</u> that the foreman, who had supervised the job, was responsible and fired him immediately.

(d) After investigating the accident, the manager <u>believed</u> the foreman, who had supervised the job, was responsible and fired him immediately.

On the first hypothesis suggested above, phoneme monitoring should be slower for (24b) than for either (24a) or (24c) for a target located within the relative clause, e.g., the /s/ in <u>supervised</u>. Similarly, paraphrasing should be less accurate for (24b) than for (24a). (No comparison with the paraphrasing data for (24c) is possible; the structure differs too radically.)

completion and sentence verification experiments. This might result from the decision point being determined on different bases for the two kinds of decisions. We have argued earlier that there is considerable reason for suspecting that decisions about the structure are not made until after the boundary marking the end of that clause has been passed. Thus, the additional load caused by such an ambiguity should be detectable from the point at which the hearer discovers that multiple structures are possible until some point shortly after the clause boundary when a choice between structures is made. It seems unlikely that lexical ambiguities should be sensitive to such clause boundaries, for there is no greater reason to believe that for ambiguities involving only the senses of a lexical item disambiguating information will become available within the clause than that it will become available after the end of the clause. Thus, it may be that in the absence of disambiguating context whether a decision is made between readings is more a function of the time elapsed since the ambiguity occurred than of whether or not the end of the clause has been reached.

It is interesting in this context to note that the lexically and structurally ambiguous sentences used by Foss differed in a relevant way. For nearly all of the lexically ambiguous sentences, the monitoring target occurred within the same clause as the ambiguity. However, for the structurally ambiguous sentences, the target usually occurred immediately following the end of the clause (e.g., the /b/ in the conjuction but joining the clauses). Unfortunately, since the existence of a clause boundary was nearly completely confounded with ambiguity type, there is no way of determining from Foss' data whether clause boundaries are important.

An unpublished experiment by Hakes and Foss was designed to shed light on this question. The experiment used sets of lexically ambiguous (e.g., (39)) and structurally ambiguous (e.g., (40)) sentences in which both the distance between the ambiguity and the monitoring target and

(39) (a) When the supervisor discovered that the worker had damaged the valuable plane, he fired him on the spot.
(b) When the supervisor discovered that the worker had damaged the valuable plane through his clumsiness, he fired him on the spot.

(40) (a) The shooting of the sergeant was a complete surprise to his superior officers and to the troops.
(b) The shooting of the sergeant during target practice was a complete surprise to his superior officers and to the troops.

whether the two occurred in the same or different clauses were systematically varied. Sentence (39) contains the ambiguous word plane. In (39a), this is followed immediately by a clause boundary, and the monitoring target for this sentence was the /f/ in fired. In (39b) an adverbial phrase has been added, increasing the distance between the ambiguity and the clause boundary. Two different monitoring targets were

used for these cases with lengthened clauses: the /f/ in fired, which is across the clause boundary and also a considerable distance from the ambiguity, and the /c/ in clumsiness, which is within the same clause but about the same distance from the ambiguity as the target in (39a). Thus we have three target location conditions: within the clause and close; out of the clause and close; and out of the clause and distant. Two control sentences were constructed for each of these versions of the ambiguous sentence, e.g., one replacing plane with airplane, the other replacing it with saw, yielding a total of nine conditions.

Similarly, nine conditions were created for the structurally ambiguous sentences, involving three combinations of clause boundary and monitoring target location and also involving one ambiguous and two control versions of each sentence. For the example presented in (40), the ambiguity of whether someone shot someone (or something) was resolved in the control versions by replacing shooting with murder and performance.

Eighteen sets of sentences of each ambiguity type were constructed and presented to nine groups of 10 subjects, each group receiving two sentences in each of the nine conditions for each ambiguity type. Unlike most of our phoneme monitoring experiments, subjects here were not required to paraphrase the sentences. They were warned, however, that occasionally during presentation of the sentences they would be asked to write out a paraphrase of the last sentence heard. All the sentences for which paraphrases were required were filler sentences, unrelated to the ambiguous and control sentences.

The results may, unfortunately, be described very simply. For neither ambiguity type did any effect of interest approach significance. This holds not only for effects of clause boundaries and of distance between the ambiguity and monitoring target, it also holds for the ambiguity effect itself. That is, under no conditions were the latencies for the ambiguous sentences different from those for their controls in any interesting way, thus failing to replicate earlier results. There was, for example, a significant main effect for the ambiguity variable for the structural ambiguities. The effect is hardly interesting, however, for the latencies for the ambiguous sentences were intermediate to those for the two control conditions.

It is, however, not clear that these results should be taken very seriously. Because of the great difficulty of constructing sentences which could be used in all nine conditions of the experiment, only 18 sentences of each ambiguity type were used. Given the necessity of using nine conditions, this resulted in each subject receiving only two sentences representing each condition. Consequently, the latencies entering into the analyses are highly confounded with subject effects and are considerably less stable than those we usually obtain.[18] The same difficulty in constructing sentences resulted in many of the sentences used

[18]We generally attempt to obtain latencies on at least five sentences in any given condition from each subject.

being considerably less felicitous than we would have liked, and this may be another source of the unusually high within-condition variability.

Sentence infelicity may also have contributed to another difficulty we encountered. Generally, we have discarded data from subjects in phoneme monitoring experiments who fail to respond to more than a few of the target phonemes or who produce large numbers of long latencies, e.g., longer than 1.5 sec. Usually, these criteria lead us to reject data from some 5% of the subjects tested. In this case, however, the rejection criteria resulted in our rejecting 163 subjects while using data from only 90. Clearly, something was very different for this experiment from any other we have performed.

Given the results of this experiment, we would like to believe that the results should not be taken seriously. And, as we have suggested, there are numerous arguments for doing so. Whether this is the appropriate move to make will have to be determined by further research.

A second group of experiments dealing with the relation between the processing load created by ambiguity and clause boundaries has used a variant of the probe latency task. In its usual application, the probe task involves presenting a sentence auditorily; shortly after the sentence ends, the subject sees or hears a word and is required to judge as quickly as possible whether that word occurred in the sentence. Using this task Caplan (1972) found that probe latency was shorter when the probe target was in the sentence's last clause than when it was in the penultimate clause. Caplan argued that during a clause the words in that clause are in an active processing memory and that at the end of the clause, when processing is completed, the material is transferred to another, less accessible memory. That is, in trying to find a match for the probe word, the subject first searches the material he is actively processing and only on failing to find the word there turns to searching material that has already been processed, i.e., material from earlier clauses.

On the basis of Rubenstein, Lewis and Rubenstein's (1971) data, it appears likely that an ambiguous word has multiple representations in the mental dictionary. If all of these are retrieved during comprehension processing, then it should take less time to match a probe with a target if the target is ambiguous than if it is not. But if processing of the words in a clause is completed at the end of the clause, the prediction of an ambiguity-control difference holds only if the probe is presented during the clause containing the ambiguous word. After the end of the clause, there should be no difference.

To test this prediction, Caplan (1971) developed a variant of the probe task in which the probe occurred during rather than after the sentence. Caplan then constructed a set of lexically ambiguous and control sentences such that the probe could occur a fixed distance from the target but before the end of the clause for some sentences and

after the end of the clause for others.[19] The probes were presented visually.

Caplan's results are a bit difficult to describe. Overall, the data yielded the predicted interaction. When the probe occurred within the same clause as the target, latency was significantly shorter for the ambiguous sentences than for their controls. When the probe occurred after the clause boundary, there was no difference. What renders the interpretation of these data uncertain is that the entire interaction effect is attributable to the comparison of the ambiguity with the control for one of its readings; the comparison with the control for the other reading yields no trace of an interaction. This asymmetry does not appear to be compatible with an account which holds that ambiguities are responded to more quickly because all of their readings are available during the clause.[20]

Caplan also performed a second ambiguity-clause boundary experiment using similar sentences and auditory probes. The results, while differing in detail, were generally similar to those for the visual probe study. That is, there was a trend in the data toward an interaction between ambiguity and the presence or absence of a clause boundary. The trend did not, however, reach significance.

One possible reason for the pattern of Caplan's results is that he used different sentences involving different ambiguous and control words, for the conditions in which the probe preceded and followed the clause boundary. Thus, any clause boundary effects were confounded with the effects of individual sentences and probe words. Since these are both probably sources of within-condition variance, Caplan's experiments

---

[19] Caplan's design also included a disambiguation variable such that a disambiguating context always followed the ambiguous word, varying in distance from that word such that the context could precede or follow the point at which the on-line probe was presented and could also precede or follow the clause boundary.

[20] Although Caplan makes nothing of the fact, inspection of his data reveals that there is no indication of the general clause boundary effect obtained in his other experiment. That is, although there is some evidence of an ambiguity x clause boundary interaction, reaction times for probes occurring after the clause boundary are not generally longer than those occurring within the clause. This suggests that the clause boundary effect which Caplan did obtain may be attributable to the fact that the probe occurred after the sentence was completed, i.e., the effect may be an end-of-sentence effect rather than an end-of-clause effect. If availability of material in a sentence was less after the end of the sentence but not after the end of the clause containing that material, this might also provide a part of an account of the results of the sentence verifi-

may simply have been too insensitive to detect the predicted interaction.

Hakes (unpublished)has recently completed a similar experiment using the on-line visual probe with lexically ambiguous sentences. The sentences used were taken from the Hakes and Foss ambiguity monitoring experiment described above, e.g., (39). Recall that for these sentences the distance between the ambiguous word and the end of the clause was manipulated by adding an adverbial phrase. Thus, the probe always occurred three syllables after the ambiguous target; when the sentence contained the adverbial phrase, the probe was within the clause, but when the adverbial phrase was absent, the probe followed the clause boundary. The combination of two probe locations with the ambiguity variable (the ambiguity plus controls for both readings) yielded six conditions. Three sentences of each of the six types were presented to groups of 10 subjects each.

Overall, the ambiguous probes yielded significantly faster latencies than their controls. However, the interaction of this effect with the presence or absence of a clause boundary did not even approach significance. Thus, when the probe occurred within the same clause, the mean latency was 42 ms. faster for the ambiguous probes than for their controls (averaged across the controls for the two readings). When the probe occurred after the end of the clause, the difference was 36 ms. In addition, there was a significant overall clause boundary effect. But unlike the effect reported by Caplan using the post-sentence probe, latencies were shorter following the clause boundary than preceding it.

The interpretation of the ambiguity effect here is somewhat clouded by the fact that the words used as probes (and, hence, also as targets) were different for the ambiguous and control sentences. Thus it is possible that the obtained effect is attributable to differences in processing the words themselves. It may be, for example, that the ambiguous words had a higher frequency of occurrence than their controls, a difference shown by Rubenstein et al (1971) to affect latency in a word recognition task.

We may also note in passing that the absence of an ambiguity x clause boundary interaction may have resulted from an unintended artifact of the design. We suggested earlier that in general the hearer cannot discover the boundary marking the end of a clause until he has processed material past that point (see Hakes, 1971b) for the rationale for this claim). There may, however, be cases in which the character of material within the clause permits the hearer to predict where the boundary will occur before he reaches it. This results from the fact that once the verb in a clause has been passed there are constraints on the kinds of constituents that can occur in the predicate of that clause and also on the order in which they can occur. One such constraint seems to be that a predicate adverbial (e.g., a simple adverb, a prepositional phrase or other non-clausal adverbial) cannot be separated from the clause boundary by anything other than another adverbial. Thus if the hearer notes the occurrence of a verb and then the occurrence of something analyzable as an adverbial (e.g., the beginning of a prepositional phrase), he may not need to wait until actually reaching the

end of the clause to begin assigning structure and lexical readings to the material within the clause, for the internal structure of the adverbial as well as the number of such adverbials does not affect the structure of the rest of the clause.

Unfortunately, for all of the sentences used in this experiment in which the probe occurred within the same clause as the ambiguity it occurred in a clause final adverbial phrase. So it is possible that at the point at which the within-clause probe occurred the hearer was engaging in the kinds of processing normally to be found only after the clause boundary. Clearly, the problem warrants further investigation.

Before leaving the subject of probe latency experiments, we should perhaps mention an additional experiment, recently completed, which bears on the question of the availability of information about lexical items and clause boundaries. If Caplan's claim is correct that after the end of a clause interpretations have been assigned to that clause's lexical items, it seems likely that the form of this information is basically "semantic." Clearly, early in processing a phonological representation must be available, and Crowder and Morton's (1969) results suggest that very early in processing an acoustic representation is available. It seems likely that as successively more abstract representations are constructed, less abstract ones are erased, though whether this is an automatic process or one under the hearer's control is open to question.

Concerning "semantic" and "phonological" representations, Sachs (1967) has demonstrated that within a few seconds of the end of a sentence it is probably only semantic information that is retained. Sachs presented subjects with continuous discourse which, at some predetermined point, stopped. The subject was then presented a sentence and required to judge whether the sentence was the same as one occurring in the discourse. The "probe" sentence was either identical to one in the discourse or was changed semantically, structurally or lexically, (e.g., rich substituted for wealthy). The interval between the sentence in the discourse and the test sentence was varied in terms of whether the sentence ended the discourse or was followed by additional sentences. The relevant result is that for lexical changes (i.e., synonyms) subjects performed at chance at an interval of 7.5 sec. and were only 65% correct at an interval of 3 sec. Since the 3-sec. interval was unfilled, it is difficult to tell whether performance would have been still worse had the interval been filled.

Thus, there is some reason to believe that information about a word's phonological characteristics is not retained for very long. If loss of this information is part of the clause-boundary effect postulated by Caplan, then it follows that a match of a probe and a target word must be based on semantic characteristics of the probe and target if the probe comes after the end of the clause. Thus, if the probe is a synonym of some word in a sentence it should be more difficult to reject (i.e., to say that it did not occur) than a control probe that is unrelated to any word in the sentence. This effect should be larger if the probe comes after the end of the clause containing the target. For by hypothesis, at that point only semantic information is available. There may, of course, be some difference even if the probe occurs within the same

clause if the attempt to find a match for the probe is based on both semantic and phonological information.

Hakes (unpublished) tested this hypothesis in a probe latency experiment using sentences like (41) and (42). The target in this case

(41) The student made a great many careless errors, and as a result he failed the math test.

(42) The student made a great many careless errors in his computations, and as a result he failed the math test.

was errors, the synonym probe was mistakes, and the control probe was minutes. The visual probe was presented, for (41), simultaneous with the /r/ in result, after the end of the clause containing the target. For (42), the probe was simultaneous with the /p/ in computations and within the clause containing the target. The distance between the target and the probe was equated across sentences in number of syllables.

The hypothesis received no support from the data. That is, the interaction between presence of a clause boundary and the character of the probe did not approach significance. The only effect to reach significance was the main effect for the presence of the clause boundary. And, again, this effect was opposite in direction to that reported by Caplan. Here, latency was shorter when the probe occurred after the end of the clause than when it occurred within the clause.

It might seem that the reversal of the clause boundary effect is attributable to the fact that the probes in this experiment were "out-probes", i.e., the correct response was "no" rather than "yes". That this is not the case is suggested by the fact that the clause boundary effect obtained by Hakes in the ambiguity probe experiment described earlier was in the same direction as the effect for the synonym experiment, and there the probes were "in-probes". In addition it seems unlikely that the fact of whether the probe is in the sentence or not materially affects the basic process by which the subject attempts to determine the correct answer, though it does seem likely that discovering that the probe is in fact not in the sentence may involve processing in addition to that necessary when the probe is in fact present. This is consistent with the fact that the latencies obtained in the synonym experiment were consistently longer than those obtained in the ambiguity experiment.

It seems more likely that the reason Caplan obtained different clause boundary effects than those obtained here is that his probes came after the end of the sentence. For it may well be that processing a probe when there is not also sentential material to be processed simultaneously leads to different results. The fact that in these experiments latency was shorter after the clause boundary than before may result from the fact that the post-boundary probes coincide with the early portions of a clause while the pre-boundary probe coincide with the material late in a clause. In this sense, the latency difference may simply reflect the amount of other processing that is occurring at the time the probe is presented. If this is the case, then the on-

line probe task may be well suited to determining local processing load effects within clauses, a job which the phoneme monitoring task appears to be less than ideal.

But the major question remains. What light do the result shed on the processing of ambiguous sentences and of sentences in general? The answer is, of course, that the picture is still extremely muddled. One thing which does seem to be clear is that the occurrence of an ambiguous word or a structurally ambiguous sequence of words does increase processing load. It seems fair to conclude that during comprehension processing more than one reading of an ambiguous word is retrieved and considered. Similarly, the increased monitoring latency obtained following structural ambiguities suggests that multiple hypotheses as to the correct structure to assign are considered.

However, if this is the case, we are faced with a paradox. For, as we have noted earlier, verb complexity appears to be just a case of local ambiguity. Yet while the monitoring task has yielded the expected effects for ambiguities, it has consistently failed to do so in the verb complexity experiments. The only apparent difference between the verb cases and other cases of ambiguity is that for the former the ambiguity is strictly local. The structural uncertainty created by a complex verb must, with very few exceptions, be resolved before the sentence ends and is generally resolved before the clause ends. While it seems intuitively implausible that this difference should have a dramatic effect on comprehension, it may just be that this is the cause of the consistent difference.

But beyond saying that ambiguity does increase processing load, there is little that can be said with any confidence. The results to data are simply unclear as to exactly where these effects occur, what might cause them to disappear, etc. There appears to be some reason to believe that clause boundaries are somehow implicated, but even here the question of how remains a mystery.

Before leaving ambiguous sentences, it seems appropriate to raise a related question, the question of the effects of context on the processing of ambiguities. As has often been pointed out, most content words are ambiguous when taken in isolation. Generally, however, they are not when they occur in sentential context. Thus it seems appropriate to distinguish between a word that is functionally ambiguous and one which is not. The question is, how do these cases differ?

To begin, we may distinguish two cases in which context serves to functionally disambiguate a potentially ambiguous word. One is the case in which the context preceding the word is neutral between its readings, and the context which disambiguates it comes only after the word itself. Here it seems reasonable to suppose that initially all readings for the word are retrieved (as would be the case for a word that was never disambiguated) and that processing of later material served to eliminate all but one reading. This appears to be parallel to the general case we have discussed of formulating and testing multiple hypotheses, with the difference that the case discussed involved structural hypotheses while this case involves hypotheses about the

readings of an individual word.[21] At present, of course, little of substance can be said about the nature of this process except that it appears likely that posterior context within the clause has a different effect from that following the clause. That is, if it is the case that a reading is assigned at the end of the clause regardless of whether disambiguating context has occurred, than any disambiguation later than that must be used to revise the assigned reading rather than entering into its initial determination.

To date, little research has concerned itself with the question of post-disambiguation. The sentence verification experiments described earlier could be seen as speaking to this question. But they apparently speak primarly to the question of revising an assigned reading rather than that of constraining the reading initially. The ambiguity probe experiments by Caplan (1971), discussed earlier, also provide some evidence on this question. Recall that Caplan's experiments involved lexically ambiguous and control sentences for which the probe occurred either within the same clause as the ambiguity or after the end of that clause. In addition, Caplan provided a disambiguating context which occurred within the same clause but after the ambiguous word or occurred after the end of the clause. Combining these conditions, Caplan thus had cases where the probe occurred either before or after the disambiguating context, either in or after the clause. His results suggest that, in general, the disambiguating context reduces the difference in probe latency between the ambiguous and control words. The data are unclear, however, as to whether this effect occurs only when the disambiguation occurs within the same clause as the ambiguity. One experiment (the auditory probe) suggests that this is the case; the other (visual probe) suggests that the disambiguation effect is independent of the presence of a clause boundary. For reasons mentioned earlier, these results must be interpreted with caution.

The other case in which context may functionally disambiguate a potentially ambiguous word is the case in which the disambiguating context precedes the ambiguous word. This is the more interesting case in that two quite distinct possibilities exist for how such prior context might affect interpretation of the ambiguous word. One is that prior context affects processing in much the same manner as posterior context. Such might be the case if, for example, during a clause little more processing was accomplished than merely retrieving dictionary information about the lexical items, the major work of assigning readings and integrating them into the relevant structure being deferred until the end of the clause. More generally, such a process would involve retrieving as much information about the ambiguous item when it was pre-disambiguated as when it was not.

The other possibility, one suggested by information-processing

---

[21] That there may not be a sharp separation of these cases is suggested by the fact, noted earlier, that in many cases there is a correlation between the appropriate structure and the appropriate semantic reading.

models such as those of Norman (1969) and Morton (1970), is that prior context serves to reduce the amount of material examined and retrieved from the dictionary such that readings not compatible with the prior context are not even considered. It seems intuitively likely that some such selection occurs. But the question remains of how much prior selection occurs and how. It seems entirely possible, for example, that prior context is used to select readings which have form class assignments appropriate to the context but does not select among readings that differ only semantically. As we have suggested earlier, at least some structural decisions must be made within the clause, i.e., at least those necessary for determining at least a sequence of form class assignments. Thus, it seems possible that prior context is used to select readings on this basis without at the same time selecting on the basis of the semantic content of different readings.

The earliest relevant experiment was reported by Garrett (1965), who studied the accuracy with which a click was located in sentences containing ambiguous words. The click was located in the ambiguous word, and the sentences were constructed such that a disambiguating context either preceded or followed the ambiguity, e.g., (43), where the ambi-

(43) (a) It was John's firmly held conviction that in a light car one can be seen better at night.

(43) (b) That one can be seen better at night in a light car was John's firmly held conviction.

guity of interest is light. The subjects' task was to listen to the sentence and then write it down and indicate where the click occurred. The results indicated that the clicks were located more accurately when the disambiguating context preceded the ambiguous word than when it followed. Garrett argued that pre-disambiguation served to reduce the amount of processing needed for the ambiguous word, making more processing capacity available for locating the click.

A more directly relevant experiment has been reported by Foss and Jenkins (1972), using the phoneme monitoring task. Foss used sets of sentences containing either an ambiguous word or an unambiguous control word and also containing a context prior to the ambiguity that was either disambiguating or neutral, e.g., (44).

(44) (a) The cattleman purchased the stock before the price went up.

(b) The man purchased the stock before the price went up.

(c) The cattleman purchased the cattle before the price went up.

(d) The man purchased the cattle before the price went up.

(44a) represents the case of an ambiguous word (stock) preceded by a disambiguating context (cattleman). (44b) contains the same ambiguity with a neutral context while (44c) and (44d) contain an unambiguous control word, together with the disambiguating and neutral contexts. The same set contained sentences with both a control word (securities)

and a disambiguating context word (financier) for the other reading of stock.

Foss and Jenkins' results replicated the ambiguity effect found earlier. That is, with the neutral context monitoring latency was longer for the ambiguous words than for their controls. However, the disambiguating context did not reduce this difference, indicating no pre-disambiguation effect.

On first consideration, these results would seem to indicate that prior context has a disambiguating effect only after information about the ambiguous word has been retrieved, i.e., that it serves, like post-disambiguating context, to govern the selection of a reading from the set of possible readings retrieved rather than restricting the set considered. However, inspection of the sentences used suggests another possibility. For most of the sentences, the disambiguating context occurred within the same clause as the ambiguous word. If we suppose that processing of the context itself has not been completed by the time dictionary information about the ambiguous word is retrieved, then clearly the context could not affect that retrieval. In support of this account, a re-analysis of Foss and Jenkins' data showed that for sentences where the disambiguating context was in a prior clause, the predicted reduction in monitoring latency occurred. The results of this analysis should, of course, be interpreted with considerable caution. Very small numbers of sentences entered into some of the cells of the analysis, and there were different sentences represented in different cells. Nonetheless, the results of the re-analysis suggest that the effect of a pre-disambiguating context is to reduce the amount of information considered about the ambiguous word, at least under some circumstances. If this is the case, the question remains of determining what those circumstances are.

On the assumption that processing of material in a clause is not completed until the end of that clause, it seems possible that a disambiguating context must be in a prior clause to be effective. That is, the context cannot be used to disambiguate anything until its own interpretation is completed. On the other hand, it may not be a matter of clausal structure at all but rather that there must be sufficient distance between the context. Again, such questions clearly call for further research.[22]

In summary, the results of recent research on ambiguity leave matters in a very unsatisfying situation. That ambiguity affects

---

[22]David Swinney, working in our laboratory, is currently designing a series of experiments for exploring these questions. To date, only the data from a pilot study are available, and the ambiguity effects in these appear to be considerably confounded by the fact that the ambiguous words used were of considerably higher frequency of occurrence than their controls. For this reason, and because the study is just a pilot study, we will not discuss the results here.

comprehension processing load seems indubitable. But beyond that, very little else is clear. Having nothing to suggest in the way of conclusions, we will content ourselves with suggesting why research on ambiguity has so far been so unilluminating.

One reason is a methodological one. Research on ambiguity, like research on comprehension processing in general, has been forced to use techniques whose properties are as yet largely unknown. It goes almost without saying that the picture of comprehension that emerges from any experiment is as much a function of the task used as it is of comprehension. In a sense, a theory of comprehension processing will have to be also a theory of the tasks used. Thus, for example, we have little real idea of how a task like the phoneme monitoring task reflects comprehension processing.

Foss (e.g., 1969; Foss & Lynch, 1968) originally characterized it as reflecting decision processes occurring during comprehension. The basic assumption was that the number of other decisions (e.g., about readings for lexical items, syntax, etc.) that were being made at the time the monitoring target occurred would be reflected in the time taken to make the decision that the target had occurred and to respond to it. This characterization was, of course, completely ad hoc, there being no evidence at that time for believing that it was that aspect of comprehension processing rather than some other that affected monitoring latency. And the data that have become available since that time have done little to clarify the issue.

Many of the data on ambiguous sentences could, for example, be handled quite well by a model which postulated that monitoring latency effects reflect primarily the amount of material being held in short-term memory pending processing. And it is equally possible that both active processing load and storage load contribute to monitoring latency. What is clear, however, is that we do not as yet have anything approaching an adequate model of the processing involved in the task. And this, coupled with the obvious inadequacies of our current models of comprehension processing leaves us in the predicament of not knowing, when a prediction is not confirmed, why it was not confirmed.

The situation with respect to the probe latency task is, if anything, worse, for as yet we have even fewer data available to constrain accounts of how this task reflects comprehension processing. It seems reasonable to assume that the observed effects reflect characteristics of the probe word itself and also characteristics of what else is going on at the time the probe is processed. But beyond this, anything that might be suggested would be pure, unbridled speculation.

A second problem with research on ambiguity is one even less likely of solution than the methodological problem. We argued earlier that the design features of natural languages are such as to minimize ambiguity. The very rarity of the garden-path phenomenon suggests that both the language itself and the design of the device which comprehends it operate to minimize the problem. Although it appears that comprehen-

sion is an uncertain process, it also appears that the processing device places its bets in such a way that it rarely loses. Undoubtedly, when it is forced to make a decision in the face of uncertainty, its decision is not an unbiased one; somehow it is able to take into account the differing likelihoods of the different decisions possible and thereby maximize its success.

But all of this is by way of saying that sentences that are truly, psychologically ambiguous are rare and unusual (though from a linguistic standpoint they are far more frequent). Their rarity makes the task of constructing such sentences for experimental purposes extremely difficult. And this, in turn, almost inevitably results in other problems. We have noted in connection with one of the ambiguity experiments described that the sentences used were, putting it rather charitably, infelicitous. And this seems to be characteristic of most of the sentences used in most ambiguity experiments. At present we have little knowledge of how such infelicity affects comprehension processing. But it does not seem overly pessimistic to suspect that it involves at least some distortion of normal processing. And to the extent that this is the case, we should perhaps not expect that the results we obtain will be anything other than cans of worms.

All this is not intended to force the conclusion that the task of understanding how people understand ambiguous sentences is impossible, thought it might not be irrational to so conclude. Rather, the moral seems to be that we need to take much greater care in constructing sentential material for our experiments than we customarily have been. Hopefully this, coupled with sufficient research and theorizing to enable us to better understand the tasks with which we attempt to study ambiguity, will eventually result in a far clear picture emerging than has to date.

V. Coda

Throughout this paper we have advocated the view that the process of comprehending sentences is an active process, that on-line processing of considerable complexity occurs while a sentence is being heard and that an inevitable consequence of this active processing is that it is uncertain. We have discussed a considerable body of research which has attempted to clarify some of the characteristics of this processing, of the kinds of information it utilizes and how it utilizes them. There is, of course, a much larger body of research which we have ignored. The reasons are varied, but one of the primary ones is that the relevance of much of this research is at best unclear.

There is no clear a priori basis for deciding whether data collected using a given task are relevant to comprehension processing or to some other aspect of psycholinguistic performance. Nonetheless, on empirical grounds, it seems possible to argue that many experiments most likely reflect primarily other processes. And others, it would appear, bear more on questions of what a listener can do under other

than normal circumstances than on questions of what he is likely to do normally. Thus, for example, the pattern of recall errors reported by Mehler (1963) for actives, passives, negatives and interrogatives in free-recall learning appear to have been influenced considerably by the development of recall strategies which are probably fairly specific to that task (see Bregman & Strasberg 1968).

In general, we suggest that results obtained using sentence recall (i.e., "parroting") probably cannot be taken as reflecting very directly on comprehension processing. Failure to recall correctly does not imply that the sentence was not comprehend. As we have suggested earlier, superficial analyses of a sentence (e.g., the ordered string of lexical items) tend not to be retained for very long. Yet correct recall depends on being able to reproduce that string. Thus, failure to recall verbatim may indicate failure to retain the sentences's superficial characteristics rather than failure to comprehend.

Similarly, correct recall does not entail comprehension. Cases of children reproducing verbatim nursery rhymes involving archaic words and syntax are too well-known to require extensive comment. More directly relevant is an unpublished experiment by Savin (1966) in which he reports immediate recall of doubly-self-embedded sentences with 85%-90% accuracy in the short-term memory paradigm developed by Savin and Perchonock (1965). Since experiments using other techniques (e.g., paraphrasing) rarely yield estimates of comprehension of such sentences over 60%, the most likely conclusion is that Savin's subjects were recalling the sentences without comprehending them.

Similar arguments could be constructed for many of the other techniques commonly used in psycholinguistic experiments, but it seems relatively pointless to do so. In part, the problem is that without a well-developed theory of comprehension processing there is no firm basis on which to evaluate the relevance of techniques. As we have indicated earlier, such a theory does not yet exist. Consequently, the way in which particular tasks reflect comprehension processing is still a very open question. And, as our earlier discussion of the phoneme monitoring task suggests, the relations between tasks and comprehension are likely to be very complex. To reiterate a point made earlier, the relations require much more extensive attention than they have received to date.

Undoubtedly, our lack of knowledge about how performance tasks reflect comprehension contributes a considerable amount to our uncertainty about how to evaluate the current status of the heuristic strategy model of comprehension processing. It seems clear that at this point the only fair conclusion is that the success of this model is less than overwhelming. More experiments yield negative evidence or unexpectedly complicated results than yield clear support for the model. But the question remains as to whether the fault lies primarily with the tasks used to assess comprehension or with the model itself. And even here there are two possibilities, for the model may simply be wrong in its basic approach or it may be that the model is itself not yet sufficiently well-developed to provide an adequate account of comprehension.

The optimistic conclusion would be that at this stage of conceptual and empirical development such a muddled picture is inevitable. For whatever else may be said, it is clear that the heuristic strategy model we have espoused is hardly a well-developed theory. At best it consists of a set of suggestions about the lines along which such a theory might be developed. And since, on any account, the phenomena for which the theory hopes to account are extraordinarily complex, one should perhaps not expect more than has been delivered.

It would perhaps seem worthwhile at this point to raise the question of whether there are alternative models which might serve as a better foundation for a theory of comprehension than the heuristic strategy model. The answer, unfortunately, is that if there are such models they are presently even less well developed than the heuristic strategy model. The alternative model usually suggested in this context is the DTC model which assumes, in one form or another, that comprehension processing directly reflects the rules and structures of a transformational grammar. But although many psycholinguists have talked about such models, none has as yet made a serious attempt to develop such a model (see Gough, 1971). And the conceptual and empirical problems such a model would encounter do not make such an attempt appear a particularly promising one. So it seems that we have no alternative but to hope that out of the current mare's nest of conflicting and uninterpretable results some order will eventually emerge. That this is the case is, at best, frustrating. To expect it to be otherwise would be naive.

References

Bever, T.G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.) Cognition and the development of language. New York: Wiley, 1970, pp. 279-362.

Bever, T.G., Garrett, M., & Hurtig, R. The interaction of perceptual processes and ambiguous sentences. In preparation.

Bever, T.G., Lackner, J.R., & Kirk, R. The underlying structures of sentences are the primary units of immediate speech processing. Perception and Psychophysics, 1969, 5, 225-234.

Bever, T.G., & Langendoen, D.T. A dynamic model of the evolution of language. Linguistic Inquiry, 1971, 2, 433-464.

Blumenthal, A.L. Observations with self-embedded sentences. Psychonomic Science, 1966, 6, 453-454.

Bregman, A.S., & Strasberg, R. Memory for the syntactic form of sentences. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 396-403.

Cairns, H.S. Ambiguous sentence processing. Unpublished Ph.D. dissertation, The University of Texas at Austin, 1970.

Cairns, H.S., & Foss, D.J. Falsification of the hypothesis that word frequency is a unified variable in sentence processing. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 41-43.

Caplan, D.N. Probe tests and sentence perception. Unpublished Ph.D. dissertation, M.I.T., 1971.

Caplan, D. Clause boundaries and recognition latencies for words in sentences. Perception and Psychophysics, 1972, 12, 73-76.

Chomsky, N. Aspects of the theory of syntax. Cambridge: M.I.T. Press, 1965.

Crowder, R.G., & Morton, J. Precategorical acoustic storage (PAS) Perception and Psychophysics, 1969, 5, 365-373.

Fillmore, C.J. The case for case. In E. Bach & R.T. Harms (eds.) Universals in linguistic theory. New York: Holt, Rinehart & Winston, 1968. Pp. 1-90.

Fodor, J.A., Bever, T.G., & Garrett, M.F. Language structure and verbal behavior. New York: McGraw-Hill (in press).

Fodor, J.A., & Garrett, M. Some reflections on competence and performance. In J. Lyons & R.J. Wales (eds.) Psycholinguistics papers. Edinburg: Edinburgh University Press, 1966. Pp. 135-154.

Fodor, J.A., & Garrett, M. Some syntactic determinants of sentential complexity. Perception and Psychophysics, 1967, 2, 289-296.

Fodor, J.A., Garrett, M., & Bever, T.G. Some syntactic determinants of sentential complexity, II: Verb structure. Perception and Psychophysics, 1968, 3, 453-461.

Foss, D.J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 457-462.

Foss, D.J. Some effects of ambiguity upon sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 699-706.

Foss, D.J., Bever, T.G., & Silver, M. The comprehension and verification of ambiguous sentences. Perception and Psychophysics, 1968, 4, 304-306.

Foss, D.J., & Jenkins, C.M. Some effects of context on the comprehension of ambiguous sentences. (in preparation)

Foss, D.J. & Lynch, R.H., Jr. Decision processes during sentence comprehension: Effects of surface structure on decision times. Perception and Psychophysics, 1969, 5, 145-148.

Garrett, M. Syntactic structures and judgments of auditory events. Unpublished Ph.D. dissertation, University of Illinois, 1965.

Garrett, M.F. Does ambiguity complicate the perception of sentences? In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.) Advances in psycholinguistics. New York: American Elsevier, 1971. Pp. 48-60.

Gough, P.B. (Almost a decade of) experimental psycholinguistics. In W. O. Dingwall (Ed.) A survey of linguistic science. College Park: University of Maryland Press, 1971.

Hakes, D.T. Psychological processes involved in comprehending sentences. Unpublished manuscript, 1969.

Hakes, D.T. Does verb structure affect sentence comprehension? Perception and Psychophysics, 1971, 10, 229-232.(a)

Hakes, D.T. Psychological processes in sentence comprehension. Unpublished manuscript, 1971. (b)

Hakes, D.T. Effects of reducing complement constructions on sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 278-286.

Hakes, D.T., & Cairns, H.S. Sentence comprehension and relative pronouns. Perception and Psychophysics, 1970, 8, 5-8.

Hakes, D.T., & Foss, D.J. Decision processes during sentence comprehension: Effects of surface structure reconsidered. Perception and Psychophysics, 1970, 8, 413-416.

Johnson, N.F. The psychological reality of phrase-structure rules, Journal of Verbal Learning and Verbal Behavior, 1965, 4, 469-475.

Johnson, N.F. The influence of associations between elements of structured verbal responses. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 369-374.

Katz, J.J. Semantic theory. New York: Harper & Row, 1972.

Katz, J.J., & Postal, P.M. An integrated theory of linguistic descriptions. Cambridge: M.I.T. Press, 1964.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Lieberman, P. Intonation, perception and language. Cambridge: M.I.T. Press, 1967.

MacKay, D.G. To end ambiguous sentences. Perception and Psychophysics, 1966, 1, 426-436.

MacKay, D.G., & Bever, T.G. In search of ambiguity. Perception and Psychophysics, 1967, 2, 193-200.

Martin, J.E. Semantic determinants of preferred adjective order. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 697-704.

Martin, J.E. Adjective order and juncture. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 379-383.

Miller, G.A., & Chomsky, N. Finitary models of language users. In R.D. Luce, R.R. Bush & E. Galanter (eds.) Handbook of mathematical psychology, Vol. 2. New York: Wiley. Pp. 419-492.

Morton, J. A functional model for memory. In D.A. Norman (Ed.) Models of human memory. New York: Wiley, 1970.

Norman, D.A. Toward a theory of memory and attention. Psychological Review, 1968, 75, 522-536.

Rohrman, N.L. The role of syntactic structure in the recall of English nominalizations. Journal of Verbal Learning and Verbal Behavior., 1968, 7, 904-912.

Rohrman, N.L. More on the recall of nominalizations. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 534-536.

Rosenbaum, P.S. The grammar of English predicate complement constructions. Cambridge: M.I.T. Press, 1967.

Rubenstein, H., Lewis, S.S., & Rubenstein, M.A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Sachs, J.S. Recognition of semantic, syntactic and lexical changes in sentences. Paper presented at the Psychonomic Society meetings, Chicago, October, 1967.

Savin, H.B. Grammatical structure and the immediate recall of English Sentences. Unpublished manuscript, 1966.

Savin, H., & Perchonock, E. Grammatical structure and the immediate recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 348-353.

Studdert-Kennedy, M. The perception of speech. In T.A. Sebeok (Ed.) Current trends in linguistics, Vol. 12. The Hague: Mouton, 1971.

Vendler, Z. Adjectives and nominalizations. The Hague: Mouton, 1968.

Watt, W.C. On two hypotheses concerning psycholinguistics. In J.R. Hayes (Ed.) Cognition and the development of language. New York: Wiley, 1970. Pp. 137-220.